CEOS Open Search best practices document outline

CEOS Open Search Best Practices

Abstract

Introduction

Diagram of OSDD -> Search -> Response

Specification and extension adherence

Calin Duma I think here we have to go into a little bit more detail, for example the Parameter Extension is mandatory only if implementer chooses to support additional parameters. Also, if we can help the reader avoid looking at the 8 documents below, we should do it.

Doug Newman after going over the parameter extension in order to implement it in ECHO I've concluded that it is backwards compatible with the base specification. The don't have to use the Parameter element. I think we should require they do user it in order to achieve our 'programmatic client' though.

Doug Newman '8 documents' - tricky - I don't want to regurgitate information in those documents for fear of them diverging. On the other hand, it needs to be readable in isolation. My approach is to define the mandatory documents up front and then tackle them in a more verbose manner later in the document where certain rules are crucial to our project. Two-step search using OSDD links is an example of this.

Archie Warnock Listing the definitive references here makes researching details much more convenient. I'd recommend leaving them and omitting duplicate details from this document as much as possible.

- Base specification - **MANDATORY**
- Parameter extension - **MANDATORY**
- Geo extension - **MANDATORY**
- Time extension - **MANDATORY**
- Relevance extension - **OPTIONAL**
- OGC candidate specification - **MANDATORY**
- ATOM - **MANDATORY**
- ESIP Best Practices **- PARTIAL**

The Open Search Descriptor Document

Anatomy of an OSDD

Draft 2 of OSDD specification

Use name, title and value for any parameters described

Obtain OSDD by form requiring clientId - as per ESIP Best Practices

Use profiles for search term language specification

Use profiles for WKT subset of Geometry parameter if applicable

Example from ECHO Open Search implementation

The search request

Anatomy of a search

Support for search by geo box - as per ESIP Best Practices and OGC Specification

Support for search by time start and stop - as per ESIP Best Practices and OGC Specification

Support for search by free text - as per ESIP Best Practices and OGC Specification Archie Warnock - We should recommend a standard response for remote sources which do not support free text search. Doug Newman this would be handled by a lack of such

of query parameter in the OSDD no? So a request containing that query parameter would result in a BAD_REQUEST response? Archie Warnock - I guess so. Would that still constitute "Support for search by free text - as per ESIP Best Practices"? Maybe we should note that support for free text search is, of necessity, optional.

Support for result set navigation by startPage and count is mandatory

Support for Client ID parameter is mandatory - as per ESIP Best Practices Archie Warnock - By which you mean pass it through transparently. No actions are required other than that.

Examples from ECHO Open Search implementation

The search response

Anatomy of a response

Result traversal elements in feed Archie Warnock - including a definitive description of how itemsPerPage is to be interpreted. I still maintain that it ought to document the number of <entry> elements returned, regardless of what the request asked for but, regardless, it's an ambiguity that ought to be resolved explicitly. Similarly (perhaps), a definitive interpretation for the case when both startPage and startIndex appear.

Spatial extent by minimum bounding rectangle in georss box format in entry - as per ESIP Best Practices. Constituent geometries may also be rendered in their native geometries. Special case for multiple box extents.

Temporal extent by dublin core date in entry - as per ESIP Best Practices

Link mappings - rel types for data, metadata, browse, documentation and more searching - as per OGC specification

Representation of zero results

Representation of error conditions via 4xx or 5xx http codes plus descriptive, human-readable body

Relevancy - optional - as per Relevancy extension

Result ordering - optional - as per OGC specificiation

Examples from ECHO Open Search implementation

Two-step searching

'One serious hurdle to overcome in searching for data is the enormous number of data items to account for in responses, as well as the expected number of successful “hits” for a query. In ordinary web searches, the searcher is usually looking for a small number of web pages or documents. Relevance ranking typically does a good job of presenting these successful hits near the top of the returned list, followed by single point-and-click retrievals. However, when searching for Earth science data covering large time periods or spatial areas, a user will often specify a set of constraints to find an appropriate data collection together with space-time criteria for files within that data collection. Often, the precision of the data collections returned for the search is low, with many spurious hits. However, the space-time precision of the files is often quite high: that is, the user truly wants to use all the data files of a desirable data collection set that fall within the space-time region of interest. Thus, searching for all data satisfying both dataset content and space-time region at the same time can produce a great many spurious hits, i.e., all the files for data collections that are not desired.'

'To get around this precision problem CEOS Open Search Best Practices, in line with ESIP Open Search Best Practices, define a two step search process. The data collection search is performed first. The results list, rather than being specific data items, comprises data collections; each data collection includes a link to the OpenSearch Description document that describes how to search for files within that data collection. Open Search Description Documents can be recognized in the Atom response by the type attribute: type=”application/opensearchdescription+xml”.

'Typically, the data collection being searched is constrained in the search template URL through an extra parameter'

Diagram of two-step concept

Links to more searches point to OSDDs - as per ESIP Best Practices

Links to more searches append client id from original search

Examples from ECHO Open Search implementation

Example Architecture - IDN and CWIC

Architecture diagram

Client use case for IDN and CWIC implementation

Here

References

Contacts